

1) What are the movie titles that the user has rated?

Code –

```
REGISTER '/home/acadgild/Downloads/piggybank-0.15.0.jar';

M = load '/home/acadgild/Downloads/movies.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
AS (movieid:int,title:chararray,genres:chararray);

R = load '/home/acadgild/Downloads/ratings.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
AS (userid:int,movieid:int,ratings:float,timestamp:chararray);

titles_data = FOREACH M generate movieid,title;

ratings_data = FOREACH R generate userid,movieid,ratings;

user_ratings = FILTER ratings_data BY ratings is not null;

dis_movieid = FOREACH user_ratings GENERATE movieid;

join_title = JOIN dis_movieid by movieid,titles_data by movieid;

res_title =FOREACH join_title GENERATE title;

final_title = DISTINCT res_title;

store final_title into '/home/acadgild/Downloads/Pig/output5';
```

Output- showing subset of the movie list rated by the users

```
[acadgild@localhost output4]$ cat part-r-00000
Angela (1995)
Angele (1934)
Angst (1983)
Animal (1977)
Animal (2005)
Animal (2014)
Anjali (1990)
Antboy (2013)
Apache (1954)
Aprile (1998)
Ararat (2002)
Arcade (1993)
Arcana (1972)
Argo 2 (2015)
Arnold (1973)
Arthur (1981)
Arthur (2011)
As One (2012)
Ashley (2013)
```

Output-

2) How many times a movie has been rated by the user? And the average rating for a movie?

Code –

```
REGISTER '/home/acadgild/Downloads/piggybank-0.15.0.jar';

M = load '/home/acadgild/Downloads/movies.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
AS (movieid:int,title:chararray,genres:chararray);

R = load '/home/acadgild/Downloads/ratings.csv' USING
org.apache.pig.piggybank.storage.CSVExcelStorage(',', 'NO_MULTILINE', 'UNIX', 'SKIP_INPUT_HEADER')
AS (userid:int,movieid:int,ratings:float,timestamp:chararray);

group_by_movie = GROUP R BY movieid;

movie_user_count1 = FOREACH group_by_movie GENERATE group as movieid,COUNT($1) as cnt;

movie_rating_sum1 = FOREACH group_by_movie GENERATE group as movieid,SUM(R.ratings) as
sum;

movie_rating_avg1 = JOIN movie_user_count1 by movieid,movie_rating_sum1 by movieid;

movie_rating_res1 = FOREACH movie_rating_avg1 GENERATE movie_user_count1::movieid,cnt as
usercount,(float)sum/cnt as avg;

movie_user_avg_res1 = JOIN movie_rating_res1 by movieid,M by movieid;

movie_user_avg_final1 = FOREACH movie_user_avg_res1 GENERATE M::title as
moviename,usercount,avg;

store movie_user_avg_final1 into '/home/acadgild/Downloads/Pig/output3';
```

Output- showing movie rating count and average for 20 movies from the output file

```
[acadgild@localhost output3]$ cat part-r-00000 | head -20
Toy Story (1995)          66008    3.8881576
Jumanji (1995)  26060    3.2369533
Grumpier Old Men (1995)  15497    3.1755502
Waiting to Exhale (1995)    2981    2.8757129
Father of the Bride Part II (1995)    15258    3.0795648
Heat (1995)    27895    3.8417637
Sabrina (1995)    15157    3.3721054
Tom and Huck (1995)    1521    3.1291914
Sudden Death (1995)    4423    3.0083654
GoldenEye (1995)    32534    3.4318404
American President, The (1995)    19475    3.6605904
Dracula: Dead and Loving It (1995)    4475    2.6721787
Balto (1995)    1838    3.3264418
Nixon (1995)    6807    3.4308066
Cutthroat Island (1995)  3125    2.72704
Casino (1995)    20517    3.795511
Sense and Sensibility (1995)    24292    3.952227
Four Rooms (1995)    6090    3.4090312
Ace Ventura: When Nature Calls (1995)    24212    2.6339831
Money Train (1995)    4621    2.8940706
```

